Biotagging, an *in vivo* biotinylation approach for cell-type specific subcellular profiling in zebrafish.

Le A Trinh[1*], Vanessa Chong-Morrison[2] and Tatjana Sauka-Spengler[2,*]

1 University of Southern California
Department of Biological Sciences
Molecular and Computational Biology
Translational Imaging Center
Los Angeles, CA 90089

2 University of Oxford
Weatherall Institute of Molecular Medicine
Radcliffe Department of Medicine
Oxford, OX3 9DS

* Corresponding authors: letrinh@usc.edu and tatjana.sauka-spengler@imm.ox.ac.uk

Keywords: cell-type specific *in vivo* biotinylation, nuclear transcriptome, enhancer RNA

Abstract

Interrogation of gene regulatory circuits in complex organisms requires precise and robust methods to label cell-types for profiling of target proteins in a tissue-specific fashion as well as data analysis to understand interconnections within the circuits. There are several strategies for obtaining cell-type and subcellular specific genome-wide data. We have developed a methodology, termed “biotagging” that uses tissue-specific, genetically encoded components to biotinylate target proteins, enabling in depth genome-wide profiling in zebrafish. We have refined protocols to use the biotagging approach that led to enhanced isolation of coding and non-coding RNAs from ribosomes and nuclei of genetically defined cell-types. The ability to study both the actively translated and transcribed transcriptome in the same cell population, coupled to genomic accessibility assays has enabled the study of cell-type specific gene regulatory circuits in zebrafish due to the high signal-to-noise achieved via its stringent purification protocol. Here, we provide detailed methods to isolate, profile and analyze cell-type specific polyribosome and nuclear transcriptome in zebrafish.

1. Introduction

Complex multicellular organisms are comprised of numerous cell types, each executing distinct regulatory programs. The precise deployment of these regulatory regimes to express discrete set of genes through the control of gene regulatory circuits defines their cellular identity, morphology and function. Genome-wide studies of regulatory networks in specific cell populations is complicated by the fact that different cell types are often intermingled and present in small numbers within complex tissue. These issues highlight the need for approaches that allow for the efficient isolation of individual cell populations within complex biological samples.

Critical to the execution of the cell-type specific regulatory circuits is the precise cellular compartments within which the regulatory programs are carried out. Information about active, nascent transcription and mRNA processing reside within the nucleus, while translation occur in the cytoplasm. Direct comparison of whole cell, ribosomal and nuclear RNA pools show distinct differences in coding and non-coding RNA (ncRNAs) content [1-3]. The functional separation and differential content of RNA species within subcellular compartments highlight the need for genome-wide analyses of regulatory networks to take into account specific cell populations as well as cellular compartments to achieve the highest signal-to-noise information as possible. This allows regulatory circuits to be distinguished beyond cataloguing the presence or absence of transcripts.

The issue of subcellular compartmentalization of RNA is most relevant in the analysis of regulatory RNAs or ncRNAs that act as fine-tuners of the gene regulatory machinery [4]. Many regulatory RNAs are found in both specific cell-types and localized to discrete subcellular compartments. These include small nucleolar RNA (snoRNA), small interfering RNA, Piwi-interacting RNA (piRNA), small interfering RNA (siRNA), and micro RNA (miRNA) that have been the focus of extensive research over the past few decades[5-8]. In addition, the advent of next-generation sequencing technology has uncovered a host of ncRNAs that are pervasively transcribed within the nucleus[9-12]. The functions of these new classes of ncRNAs are less well established and are the focus of intense research. These include long non-coding RNA (lncRNA) (including long intergenic/intervening non-coding RNA, lincRNA), circular RNA (circRNA) and enhancer-derived RNA (eRNA) [12-20]. They are expressed dynamically and are often low in abundance [21]. These characteristics make ncRNAs challenging to study and require purification approaches that are rapid, robust and exhibit high signal-to-noise.

A number of approaches have been employed to enable cell-type specific profiling of regulatory networks in the *in vivo* context of complex tissues. Fluorescence activated cell sorting (FACS) are useful for isolating subpopulations of cells and are most widely employed when cells

can be genetically labeled with a fluorescent marker. Laser microdissection (LMD) allows isolation of cells with microscopic precision. Both of these techniques require specialized instrumentation and often involve lengthy processing times, during which cell state and gene expression can change. Furthermore, these approaches focus on isolation of the whole cell, precluding study of subcellular compartmentalization of RNAs.

Recently, strategies for *in vivo* biotinylation have been implemented to meet the challenge of cell-type specific and subcellular profiling in plants and animals [22-26]. The basic strategy involves a genetically encoded binary system that expresses the biotin ligase (BirA) and a biotin acceptor peptide (Avi-tag) fused to a protein of interest (POI). The Avi-tagged protein is biotinylated when the two components are expressed in the same cell. This approach allows the use of streptavidin-based affinity purification of the Avi-tagged protein targets. The strong streptavidin-biotin interaction with a dissociate constant of $K_d=10^{-14}$ allows for high affinity binding and thus have been widely used for purification of proteins and their interacting partners [27-30].

To permit the precise isolation of specific subcellular compartments of genetically defined cells within the zebrafish model, we have adapted *in vivo biotinylatio*n of ribosomes and nuclei, a technique that we term “biotagging”[31]. The strategy involves generating Avi-effector lines that express Avi-Rpl10a and Avi-Rangap fusion proteins to tag ribosomes and nuclei, respectively. Biotinylation of the Avi-tag fusion protein is achieved by introducing biotin ligase, BirA, into cells of interest using tissue specific BirA drivers (Fig. 1).

Here, we provide a detailed protocol of the biotagging pipeline, focusing on the nuclear purification of discrete cells within zebrafish embryos. We discuss the distinct information that subcellular purification can recover, including coding and non-coding RNAs. Additionally, we highlight quality control measures that one can employ to assess datasets gathered from the biotagging approach. Finally, we reflect on the potential of this approach for dissecting gene regulatory circuits beyond transcriptome analyses.

2. Overview of biotgging toolkit and consideration for subcellular compartment isolation

The Biotagging toolkit in zebrafish consists of two genetic components, the BirA drivers and Avi-effectors (Fig. 1) [31]. The biotinylation “driver” line expresses BirA in specific cell populations while the “effector” transgenic lines ubiquitously expresses Avi-tagged target proteins to be biotinylated. Crossing driver and effector transgenic lines results in embryos that are double transgenic. In cells where the BirA and Avi-tagged proteins are present, a single biotin moiety is added to the Avi-tagged target protein. In addition to the expression of BirA, the

expression cassette for the driver lines have the 2A peptide which allows for simultaneous expression of a fluorescent membrane-mCherry (membCherry) reporter, while the Avi-tagged effectors, have Cerulean (a blue fluorescent protein) fused to the Avi-tagged protein. The double fluorescent reporters allow for sorting and screening of embryos to ensure that cells of interest are labeled for isolation and purification. Furthermore, the fluorescent reporters enable dynamic imaging of profiled cells. Details of transgenic constructs, available lines and their expression patterns have been previously described [31]. Additionally, plasmids for generating new BirA drivers and Avi-effectors are available through NCBI and Addgene.

A number of transgenic approaches can be employed to generate the BirA drivers and Avi-effectors. BirA driver lines can be generated using Tol2-mediated plasmid or recombineered BAC transgensis, as well as more efficient maize transposon (Ac/Ds) transgenic approaches. To generate defined BirA expression patterns, the BirA expression cassette is placed under the control of enhancers and tissue- or cell- specific promoters. A caveat to using Tol2-mediated transgenesis for expression of BirA is the variation in expression patterns of different founders observed with the same expression construct[31]. Therefore, it is highly recommended that a number of different founders for each new BirA driver line should be identified and characterized using the membCherry reporter to determine the precise expression of BirA prior to use of new driver lines. However, expression levels of BirA drivers can vary across a large range (see **3.6 Experimental considerations for biotagging)** as we have previously shown that biotinylation of Avi-tag proteins is not affected by varying levels of BirA expression [31]. Rather, it is important to determine that BirA and the Avi-tag proteins are expressed in the cell-types of interest by imaging the respective fluorescent reporters, Cerulean and mCherry.

To date, there are two Avi-effector lines for the biotagging system that ubiquitously express Avi-tagged proteins localized to the ribosomal or nuclear compartments of genetically defined cells, Avi-Rpl10a (riboAvi) and Avi-Rangap (nucAvi), respectively[31]. The riboAvi line is similar to previous **T**ranslating **R**ibosome **A**ffinity **P**urification (TRAP) [32] approach in which a protein within the large ribosomal subunit, Rpl10a, is fused to the Avi-tag and Cerulean, while the nucAvi lines uses a strategy similar to the previous *in vivo* biotinylation strategy developed in plants known as **I**solation of **N**uclei **TA**gged in specific **C**ell **T**ypes (INTACT)[22]. The INTACT approach uses the WPP domain of Ran-GTPase Activating Protein 1, RanGAP1, which localizes to the nuclear pore complex of eukaryotic cells except in yeast [33, 34]. To implement a similar strategy in zebrafish, we initially used the equivalent region from the zebrafish Rangap protein and found that the Avi-tagged fusion proteins inconsistently associated with the nuclear envelope and resulted in embryonic phenotypes when injected into the embryo at high

concentrations. However, the C-terminal domain of the avian Rangap protein exhibited consistent nuclear localization. Comparison of the chicken and zebrafish cloned Rangap domains indicated that the C-terminal portion of the Rangap protein from the chicken, but not zebrafish contains a Ran Binding Domain, which associates with nuclear pore complexes [35, 36]. The difference in localization between the C-terminal Rangap domains from the two species highlight the importance of having an independent tag such as the Cerulean reporter that can be used to verify the precise localization of Avi-effectors.

Signal-to-noise achieved with cell-type specific nuclear purification:

Analysis of RNA and sequencing of cDNA libraries generated from purified nuclei based on the biotagging system indicates that the isolation of cell-type specific nuclear compartments for RNA profiling increases signal-to-noise over whole cell as it reduces the ribosomal content of isolated total RNA. Previous comparison of nuclear and whole cell RNA showed that there is a significantly smaller fraction of 18S and 28S ribosomal RNAs (rRNAs) in nuclear total RNA [37]. Isolation of cell-type specific nuclear RNA from biotagged nuclei compared to FACS shows a significantly smaller fraction of 18S and 28S rRNA (5% of total nuclear RNA vs. 50% of whole cell (FACS) total RNA) from migratory neural crest cells [31]. The low level of rRNA content yields a better dynamic range for downstream analyses in RNA-seq studies and requires less sequencing depth per cDNA library. Importantly, cDNA libraries generated from nuclear RNA for RNA-seq are enriched in nascent transcripts, bidirectionally transcribed genes and enhancer RNAs suggesting that the nuclear RNA directly reflects the regulated transcriptome [31].

The reduced levels of rRNAs is a signature of purified nuclear RNA that can be used to monitor the biotagging purification protocol (described in Materials and Methods). Bioanalyzer or TapeStation profiles of total RNA isolated from whole cells, ribosome and nuclei have distinct profiles (Fig. 2). Total RNA isolated from cells obtained by FACS or whole embryos lysis show two prominent 18S and 28S rRNA peaks that constitute the majority of cellular RNA species (Fig. 2C). Similarly, RNA profiles from biotagged and purified polysomes exhibit distinct 18S and 28S rRNA peaks reminiscent of FACS and whole embryo profiles (Fig. 2B). Conversely, nuclear total RNA show a broader range of sizes and a significantly smaller fraction of 18S and 28S rRNA peaks (Fig. 2A). This consistent broad range of RNA sizes can be used as a quality control for the complete lyses of cells in the nuclear purification protocol.

While comparison of gene content between ribosomal and nuclear RNA show significant overlap for the two compartments, there are major differences between the cellular compartments in the expression level, types of the transcripts and presence of non-coding

transcripts [31]. The vast majority of transcripts found in ribosomal and nuclear RNA dataset correspond to protein-coding genes[31]. Consistent with transcription occurring in the nuclear compartment, immature transcripts containing intronic RNA are significantly detected in nuclear RNA datasets. Additionally, reads from nuclear RNA show similar levels across the entire gene body and untranslated regions, suggesting pervasive transcription at active loci. The pervasive transcription across gene bodies is not seen in whole cell or ribosomal dataset. Furthermore, the pervasive transcription is detected in intergenic regions of the genome when nuclear RNA pools are analyzed. Moreover, purification of the cell-type specific nuclei provides higher signal-to-noise in detection of bidirectional transcripts.

Nuclei purification provide access to regulatory non-coding RNA.

The biotagging approach allows rapid purification of cell type-specific cellular compartments without enzymatic digestion or fluorescence sorting. This speed translates to a better preservation of biological states for isolation of low abundance and highly unstable RNA species compared to traditional FACS-based methods or other approaches involving lengthy processing times. Comparison of purification efficiency using FACS and biotagging purification of ribosomes and nuclear show that biotagging can recover more cells per cell-type within an embryo[31]. The increase yield and signal-to-noise achieved with cell-type specific nuclear purification provides access to regulatory non-coding RNAs such as eRNAs that can be used to bioinformatically identify cell-type specific cis-regulatory modules (CRMs).

The high signal-to noise enabled by the cell-type specific nuclear purification allows recovery of low level transcriptional events at intergenic regions that are associated with active CRM[31]. Bidirectional transcripts resulting from divergent transcription associated CRM are known hallmarks of active enhancers [17, 20, 38-40]. Coupling cell-type specific nuclear RNA-seq to ATAC-seq, in the same cell population can identify CRMs within the cell-type of interest[31].

3. Materials and Methods

Here, we describe the detailed methods to perform ribosomal and nuclear purification for any cell-types in the zebrafish that are genetically labeled by the binary biotagging system. All reagents and materials in this procedure should be RNase-free and suspended in molecular-grade, nuclease-free H_2O. We have found that complete lyses of the cell membrane is critical for achieving efficient isolation of nuclear over ribosomal RNA. A component to ensure the complete lysis of cells and minimize the presence of intact cell membranes is the use of excess

hypotonic buffer. Additionally, increasing the number of washes of the bound samples on magnetic beads is important for pure cell-type specific ribosome or nuclear RNA pools. However, there is a balance between increasing the number of washes to achieve purity and degradation of the isolated RNA.

3.1 Preparation of Streptavidin beads for RNA procedures

The streptavidin Dynabeads needs to be pre-treated with a mixture of sodium hydroxide and sodium chloride solution for use in RNA procedures as they are not suppled in a RNase-free solution (see appendix: Solution A and B). For each sample to be purified, 250µg of MyOne T1 or M-280 streptavidin-coated beads (used for ribosomal and nuclear purification protocol, respectively) are pretreated by pipetting the well-mixed bead solution into a low binding, thin walled, clear 1.7 ml microcentrifuge tube (SafeSeal, Sorenson BioScience®) and placed on a magnetic stand (DynaMagTM-2, ThermoFisher Scientific) to separate the beads from the supplied supernatant. The supernatant is discarded prior to washing the beads twice with 1mL of Solution A (DEPC-treated 0.1M NaOH, 0.05M NaCl) by mixing on a nutator for 3 minutes each time followed by placement on the magnetic stand for 3 minutes to separate beads from the supernatant. The beads are subsequently washed once in Solution B (DEPC-treated 0.01M NaCl) for 3 minutes prior to suspension in Cell Lysis Buffer or Nuclear Purification Buffer (NPB) for polyribosome or nuclei isolation, respectively.

3.2 Biotagging ribosomal purification

To purify ribosomal RNA using the biotagging system, cross transgenic birA driver expressed in the cell-of-interest to the riboAvi effector line. Confirm that the fluorescent reporters, mCherry for birA driver and Cerulean for riboAvi effector, are expressed in the cell-of interest by high-resolution laser scanning microscopy. Collect dechorinated double positive BirA (mCherry+); riboAvi(Cerulean+) embryos at the desire stage. In our experience, 300-400 embryos are needed per sample before 24 hours post-fertilization (hpf). At minimum, biological duplicate samples should be prepared for each cell-type and genotype of interest.

To lyse cells, embryos are initially placed into low binding, thin walled, clear 1.7 mL microcentrifuge tubes to be washed in Cell Lysis Buffer (see Appendix for recipe) before transferring to a 7mL Dounce homogenizer. Embryos are re-suspended in Cell Lysis Buffer at a ratio of 50 embryos per 1mL of Cell Lysis Buffer without cycloheximide. Pestle A is used to dissociate embryos by applying gentle strokes until whole tissues are no longer visible, approximately 20 strokes, in the Dounce homogenizer. Confirm complete lysis by staining cells

with 1:1 dilution of Trypan blue on microscope. Once the embryos are lysed, cycloheximide is added to the cell suspension to a final concentration of 100 µg/mL and incubated at room temperature for 15 minutes to stabilize the polyribosomes. This is followed by douncing with Pestle B for 60 strokes (3 x 20 strokes with 5 minutes pausing in between) on ice to ensure complete lyses of cells without lyses of the nucleus. Care must be taken that pestle B is not faulty and has the appropriate small clearance for efficient cell lysis.

To separate the ribosomal from the nuclear supernatant, the homogenized embryos are cleared by centrifugation at 2000g for 10 minutes at 4°C and the post-nuclear supernatant removed into clean RNase-free 1.7 microfuge tubes (1 mL supernatant per tube). The nonionic, non-denaturing detergents, IGEPAL CA-630 and 07:0 DHPC 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine are added to the supernatant to a final concentration of 1% each and mixed by inverting tubes gently 10 times. Following incubation on ice for 5 minutes, the supernatant is cleared by centrifugation at 20,000 g for 10 minutes at 4°C.

To purify polyribosomes, the post-mitochondrial supernatant is removed and added to 250 µg (2.5 x 10^8 beads) of MyOne T1 streptavidin-coated Dynabeads (1 mL supernatant per 250 µg beads) prepared for RNA procedures (see 3.1 preparation of beads) by rotating at 4^{o}C for 1 hour. The polyribosomes-beads suspension is placed onto a magnetic stand (DynaMag™-2 magnet from Invitrogen, cat. no.12321D) to remove the unbound lysate. The pelleted polyribosomes-beads are washed four times (changing tubes in between washes to minimize background) in the cold room with High Salt Buffer (see Appendix for recipe) by pooling 500 µg beads per 1 mL High Salt Buffer. After the final wash, the tubes with polyribosomes-beads are placed onto a magnetic stand to remove the High Salt Buffer. The 500 µg polyribosomes-beads pellets are immediately dissolved in 200 µL of RNA lysis buffer, incubated at RT for 10 minutes and replaced onto the magnetic stand. The RNA lysis buffer containing polyribosome-bound RNA is moved into a fresh tube and snap-frozen for future use or for immediate extraction (see RNA extraction and library preparation).

3.3 Biotagging nuclear purification

To purify nuclei using the biotagging system, the nucAvi effector line can be crossed to transgenic drivers that express BirA in the cell of interest. Embryos are sorted for double positive fluorescent expression of the mCherry and Cerulean reporters and dechorinated prior to dissociation and purification of nuclei. Typically, ~100-350 embryos are needed per nuclei isolation experiment before 24hpf. The number of embryos per experiment will depend on the cell-type of interest and developmental stage. We aim for isolation of 50-100K nuclei per sample

to obtain sufficient total nuclear RNA for cDNA library production. The lower limit of nuclei from which we have been able to isolate sufficient total nuclear RNA for whole transcriptome analysis is 30K nuclei per sample. Interestingly, a comparison of samples from 30K and 50K myocardial nuclei demonstrated remarkable reproducibility as assessed by principal component analysis and scatterplot comparison of FPKM values [31].

Cells are dissociated and lysed by first washing embryos in hypotonic Buffer H (see Appendix for recipe), supplemented with 0.01% Tricaine as an anesthetic. This initial wash step is performed to wash out any residual embryo media that can be transferred with the embryos when they are collected into microfuge tubes. Embryos are re-suspended in 1 mL/50 embryos of Buffer H and transferred to a Dounce homogenizer (7 mL, Kontes Glass Co, Vineland, NJ). Using a loose fitting pestle A, embryos are dissociated with a sequence of 20 strokes while incubated on ice for 15 minutes. Complete dissociation of embryos/tissue can be monitored by Buffer H becoming cloudy and absence of whole tissue pieces present in the buffer. If pieces of whole tissue are observed after douncing with Pestle A for the recommended strokes, continue to dounce on ice until all tissue is dissolved. The cell membranes are lysed by applying 60 strokes of tight fitting pestle B (3 x 20 strokes, pausing 5 minutes on ice after each set of 20 strokes) for 15 minutes on ice. Complete lysis of cells can be checked by diluting cells with Trypan blue at 1:1 ratio and visualizing cells on a microscope. Completely lysed cells will uptake Trypan blue. Additionally, the quality of cell lysis for each experiment can be determined by RNA extraction of the supernatant unbound supernatant after nuclei-bead purification. As a final step, purity of total nuclear RNA can be monitored with Bioanalyzer or TapeStation profiles (Fig. 2, discussed above).

Nuclei are separated from the cytoplasmic pool by centrifugation for 10 minutes at 2,000 g and re-suspended in 1 mL of NPB (see Appendix for recipe). The nuclei are bond to M-280 streptavidin-coated Dynabeads by adding NPB re-suspended nuclei to 250 µg (1.5×10^7 beads) of beads prepared for RNA procedures (see 3.1, preparation of beads) and allowed to rotation at 4^{o}C for 30 minutes. The nuclei-bead suspension is further diluted with NPB containing 0.1% Triton X-100 (NPBt) to a final volume of 20 mL for purification on a flow-based system (see below) or placed on a magnetic rack (ThermoFisher, DynaMag-15 Magnet, cat. No 12301D). The high volume of NPBt allows for increased washing and hence specificity in purification of nuclei bound beads through the flow column.

In our original purification protocol [31], we devised a flow-based setup similar to a previously published approach by Deal and Henikoff that enables a controlled flow rate of nuclei-beads across a magnetic separator [41]. The set-up consists of attaching a 1mL

micropipette tip (Gentaur Reach Pipet Tip, cat. no.24-165R or Rainin, cat. no. RT-L1000S) to a 10mL serological pipette (VWR, cat. no.89130-898 or BD Falcon, cat. no.357551) that is placed on a MiniMACS separator magnet (OctoMACS Separator, Miltenyl Biotec, cat. no.130-042-109). Additionally, we pre-treated the entire apparatus with NPB + 1% BSA for 10 minutes to coat the surface of the pipettes and avoid binding of the nuclei to the plastic.

To capture the biotinylated nuclei, the suspension of nuclei-beads is allowed to flow through the 1mL pre-coated pipette tip at a rate of ~0.75 mL/min. The flow rate is controlled by attaching a two-way stopcock (Bio-Rad, cat. no.732-8102) or a T-valve (from Biorad-low pressure fitting kit, cat. no.731-8220) to the pre-coated pipette tip with a short piece of Tygon tubing (Fisher Scientific, cat. no.14-169-1C). When flow rates are faster than 0.75 mL/min, the flow-through can be drawn up with the same serological pipette and allowed to drip through the same pipet tip on the magnetic separator to maximize recovery efficiency. This process of allowing the nuclei-bead suspension to drip pass through a pipet tip on the magnetic separator is repeated with a fresh 20mL of NPBt once more before eluting the nuclei-bound beads with 1mL of NPBt into a 1.7mL microfuge tube.

To process the purified nuclei for RNA extraction the microfuge tube containing the bound nuclei-beads is placed on a magnetic stand (ThermoFisher, DynaMag™-2 magnet cat. no.12321D) to remove the NPBt supernatant. The purified nuclei-beads can be processed for RNA extraction (see 3.4, RNA extraction and library preparation) by dissolving in 100uL of RNA lysis buffer from the RNAqueous Micro Scale Total RNA Isolation Kit (Ambion, cat. no. AM1931). The nuclei-beads are incubated at room temperature for 10 minutes and replaced onto the magnetic stand to separate out the beads from the total nuclear RNA. The RNA lysis buffer containing total nuclear RNA is then removed into a fresh tube and snap-frozen for future use or for immediate extraction and library production.

Recently, we adapted an alternative bulk purification approach using a magnetic stand that holds 15mL tubes. This alternative approach reduces the time it takes to isolate biotinylated nuclei, leading to better recovery of transcriptome states. To apply this purification scheme, the 1mL solution of nuclei-beads in NPB is re-suspended with 9mL NPBt into a 15mL Falcon tube after nuclei are bound to magnetic beads for 30minutes at 4^{o}C. The Falcon tube containing nuclei bound beads is placed on a magnetic rack that holds 15mL tubes (ThermoFisher, DynaMag-15 Magnet, cat. no.12301D). After magnetic beads are bound by the magnet for 3 minutes, the NPBt supernatant containing unbound nuclei are removed using a 1mL pipette, being careful to not disturb nuclei-bead complex that adhered to the side of the Falcon tube. The Falcon tube containing nuclei-bead is removed from the magnetic rack and nuclei-beads

are re-suspended in another 10mL of NPBt. Washing of the beads on the magnetic rack is repeated twice before purified nuclei-beads are re-suspended in 1mL ice-cold PBS and processed for RNA extraction.

3.4. RNA extraction

RNA extraction after purification of the nuclei and polyribosomes pools can be performed using the RNAqueous Micro Scale Total RNA Isolation Kit (Ambion cat. no.AM1931). After purification of ribosomes or nuclei, samples are suspended in 100uL of RNA lysis buffer from the RNAqueous RNA isolation kit. 50uL of 100% ethanol is added to the lysis buffer and mixed by pipetting or vortexing briefly to ensure complete lysis of the nuclei or dissociation of polyribosomes from beads. The solution is incubated at room temperature for 10 minutes before replacing onto the magnetic stand to separate the beads from the RNA solution. The RNA lysis buffer containing total ribosomal or nuclear RNA is removed into a fresh tube and snap-frozen for future use or for immediate extraction.

It is recommended that an additional 1.25volume of ethanol (125uL) should be added to the RNA lysis buffer to efficiently recover small RNAs such as tRNA, 5S rRNA, and micro RNA. Such a modification can be made if recovery of small RNAs is a goal of the experimental pipeline. However, we have not used such a modification and have been able to recover regulatory RNAs such as lncRNAs and eRNAs [31].

To extract RNA from the lysis buffer, the lysate is loaded onto a MicroFilter cartridge and centrifuged for 15 seconds at maximum speed or until all the mixture has passed through the filter. The flowthrough is discarded before adding 180uL of Wash Solution 1 (supplied in RNAqueous kit). Microfilter with Wash Solution 1 is centrifuged for 15seconds at maximum speed or until all the mixture has passed through the filter. This wash processed is repeated with 180uL Wash Solution 2/3 (supplied in RNAqueous kit) and centrifuged cartridge at maximum speed for 1 minute to remove residual fluid and dry the filter. RNA is eluted from the microfilter cartridge in fresh microfuge tube by applying 25uL of preheated to 75^{o}C elution solution (supplied in RNAqueous kit) to the center of filter and allowed to sit at room temperature for 1 minute to maximize elution of RNA. Microfilter cartridge is centrifuged for 1 minute at maximum speed to elute RNA from filter. Genomic DNA is removed from RNA solution with rDNaseI (provided with Ambion cat.no. AM1931) treatment for 20 minutes at 37^{o}C.

The quality of isolated RNA can be assayed using Agilent RNA 6000 Pico kit (Agilent Technologies, cat. no.5067-1513) on the Agilent 2100 Bioanalyzer, as specified by the manufacturer, or with High Sensitivity RNA Screentape (Agilent Technologies, cat. no.5067-

5579, 5067-5580) on the Agilent 2200 TapeStation. To ensure that cells are lysed and distinct subcellular compartments have been isolated, Bioanalyzer/TapeStation profiles for both experimental and lysate samples are examined for expected quantity ratios. The amount of RNA in the lysate sample should be much higher than the experimental sample where the lysate can have 10-100 fold more RNA compared to samples. This ratio is dependent on the cell-type of interest. However, the Bioanalyzer/TapeStation profiles of lysate and samples should appear similar with regards to 18S and 28S rRNA ratio and relative levels of rRNAs to other RNA species (Fig. 2).

3.5 Directional library preparation and sequencing

Isolation of subcellular compartments using *in vivo* biotinylation has been deployed in plants and a number of animal models[22, 24, 26, 41]. However, the focus has been on polyA-enrichment of RNA, thus harvesting the spliced portion of the nuclear transcriptome and RNA species processed and exported to the cytoplasm. To capture both coding and non-coding transcriptome, a directional sequencing protocol should be employed.

For directional RNA-sequencing, total RNA from nuclear and polyribosome purification protocols are first depleted of rRNA using Ribo-Zero™ Magnetic Kit (Illumina, cat. no. MRZH116), according to manufacturer's instructions. Briefly, storage liquid is removed from manufacturer supplied magnet beads by placing 225uL of beads per sample in an Eppendorf tube on a magnetic stand until all beads adhere to the walls and liquid is cleared (~1 minute). Storage liquid is removed and beads are washed twice with 225uL of RNase-free water. After washing, beads are re-suspended in 65uL of magnet bead resuspension solution and 1 µl of RiboGuard RNase Inhibitor before use in ribodepletion. For ribodepletion, approximately 50 ng of total nuclear or polyribosome RNA per library is added to a 40uL reaction that includes 4uL Ribo-Zero Reaction Buffer, 8uL Ribo-Zero Removal Solution and RNase-free water. The 40uL Ribo-Zero reaction is added to prewashed magnetic beads, mixed by vortexing and allowed to incubate at room temperature for 5 minutes before incubating at 50°C for 5 minutes to allow hybridization of rRNA to beads. Supernatant containing ribodepleted RNA is separated from the magnetic beads by placing on magnetic stand for 1 minute before transferring ribodepleted RNA to fresh Eppendorf tube.

After ribodepletion, sequencing libraries are prepared using Stranded RNA-Seq Library Preparation Kit (KAPABiosystems, Roche, cat. no. 07962142001), according to manufacturer's instructions. Ribodepleted nuclear or polyribosome RNA is re-suspended at a one-to-one volume with 2X Fragment, Prime and Elution Buffer and placed in a thermocycler to fragment

and prime for 6 minutes at 85^{o}C. Primed RNA is placed on ice before adding 10uL of 1st strand synthesis master mix. First strand synthesis is performed using the following cycle: extension for 10 minutes at 25^{o}C, synthesis for 15 minutes at 42^{o}C, and enzyme inactivation for 15 minutes at 75^{o}C. Second strand synthesis is performed by adding 30uL of 2nd strand synthesis master mix and incubating at 16^{o}C for 1hour. Stranded libraries are purified using KAPA pure beads after synthesis. Deep sequencing is performed on libraries using 50 bp paired-end reads on HiSeq2500 Illumina platform at 30-40million reads per library. We have also successfully sequenced libraries using 80 bp paired-end reads on NextSeq500 Illumina platform.

3.6 Experimental considerations for biotagging

*Balance between Avi effector and BirA levels***.**

Binding of streptavidin to biotin is highly efficient, although the total number of interactions in a given cellular space may be limited due to steric hindrance [42]. In theory, it is essential to have a critical mass of biotinylated Avi-tagged cellular compartments for efficient binding and subsequent pulldown by streptavidin. We previously demonstrated that biotinylation efficiency is sensitive to the levels of the Avi effector protein but not of BirA, highlighting that determining Avi-BirA combinations is an important point to consider when planning biotagging experiments[31]. The biotinylation reaction was shown to be rate-limited by the protein-protein interaction between biotin ligase and its substrate (in our case, the Avi-tag) [43]. This presented a possible explanation for our observation - given the size difference between the two proteins (~40 kDa BirA and ~2 kDa Avi-tag), multiple fold quantity of Avi-tag is needed for every molecule of BirA to maximize the chance of physical interaction, at least within the context of nuclei and polyribosome biotinylation by cytoplasmic BirA. Our finding that low levels of BirA are sufficient to efficiently biotinylate its target is an added strength of the biotagging system as it is not uncommon for cell type-specific markers to have low levels of expression. Therefore, BAC BirA driver lines can be preferentially generated to exploit their characteristic to express transgenes in a reproducible fashion (e.g. expressing BirA in macrophages using *mpeg2* BAC[44]).

An additional rating limiting step in the biotinylation of Avi-tag proteins is the presence of endogenous biotin in the cell type of interest. It has been previously reported that *in vivo* biotinylation in skeletal muscle cells requires biotin supplementation in the zebrafish embryo media[26]. However, biotin supplementation has not been necessary in a number of tissues in which biotagging has been performed [31, 44, 45]. The differences in endogenous levels of biotin may be due to variations in maternal diet. Therefore, an important control that ensures

the Avi-tagged proteins are biotinylated in the cell type of interest is to perform streptavidin Western blot analysis on tissue specific Avi-effectors to confirm that the Avi-fusion proteins are efficiently biotinylated.

Control datasets for comparison to sample-of-interest.

In general, two options are possible for preparing a control dataset to analyze cell-specific nuclei or polyribosome datasets. First, as we have shown previously, is to obtain a dataset from a ubiquitous BirA driver crossed to nucAvi or riboAvi effectors. Second, is to retain the unbound fraction after streptavidin bead incubation of lysates. The former approach has its strength as a closer comparison to the sample-of-interest regarding RNA quality. We have found that RNA extracted from the unbound lysate noticeably undergoes greater degradation compared to RNA extracted from the streptavidin-bound material. Economically, the first approach is advantageous as the dataset can be reused in subsequent experiments, without having to prepare and sequence different controls for every new experiment. The caveat to this approach is that the ubiquitous BirA control dataset will include transcripts from the cell type-of-interest. Depending on the abundance of the cell type-of-interest within the embryo, this may present an issue during downstream bioinformatics analyses, where differentially expressed genes may not be statistically enriched due to the transcripts-of-interest being present at high levels within the ubiquitous control. If this is a major concern, the second approach using RNA from unbound lysate should be considered.

4. Perspective

The current implementation of the biotagging enables isolation of two subcellular compartments, polyribosomes and nuclei. The versatility of the biotagging approach comes with the ability to generate new BirA drivers and Avi-effectors that can target different proteins for subcellular compartment isolation. Potential future implementation includes new Avi-effectors that target subcellular compartments such as mitochondria, ER and lysosomes. Additionally, the high binding affinity of streptavidin to biotin presents an ideal experimental system to study protein-DNA interactions with chromatin immunoprecipitation (ChIP) [46]. *In vivo* biotinylation has been employed to identify histone H3.3 occupancy in zebrafish cardiomyocytes[47]. ChIP experiments with one's favorite DNA binding protein can be performed with the creation of an Avi-fusion effector line. In addition, the use of cell-type specific nuclear dataset coupled to ATAC-seq to identify active enhancers could potentially provide new regulatory elements for the

generation of BirA drivers and Avi-effectors. Therefore, resources for expanding the biotagging toolkit can grow and refine the level of cell-type specific that can be achieved with the system.

5. Appendixes

Equipment:

Dounce homogenizer (2 or 7 mL Kontes Glass Co, Vineland, NJ)
DynaMag™-2 magnet (ThermoFisher, cat. no.12321D)
MiniMACS separator magnet (OctoMACS Separator, Miltenyl Biotec, cat. no.130-042-109)
M-280 streptavidin-coated Dynabead (ThermoFisher Scientific, cat. No. 11205D)
MyOneTM Streptavidin T1 (ThermoFisher Scientific, cat. No. 65601)
10 mL plastic serological pipette (VWR, cat. no.89130-898 or BD Falcon, cat. no.357551)
Two-way stopcock (Bio-Rad, cat. no.732-8102)
T-valve (from Biorad-low pressure fitting kit, cat. no.731-8220)
Tygon tubing (Fisher Scientific, cat. no.14-169-1C).
Low binding, thin walled, clear 1.7mL microcentrifuge tubes (SafeSeal, Sorenson BioScience®)

Reagents:

DEPC, diethyl pyrocarbonate (Sigma Aldrich, cat. no. D5758-100ML)
rRNasin (Promega, cat. no. N2111)
RNaseOUT (ThermoFisher, cat. no. 10777019)
SUPERaseIN (ThermoFisher, cat. no. AM2696)
Cycloheximide (Sigma Aldrich, cat. no. C7698-1G)
IGEPAL CA-630 (Sigma Aldrich, cat. no I8896-100ML)
cOmplete-EDTA-free protease inhibitor (Roche, cat. no.05892791001)
Ethyl 3-aminobenzoate methanesulfonate (tricaine) (Sigma Aldrich, cat. no. E10521)
DHPC, 1,2-diheptanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids 850306P-200 mg)
RNAqueous Micro Scale Total RNA Isolation Kit (Ambion cat. no.AM1931)
Agilent RNA 6000 Pico kit (Agilent Technologies, cat. no.5067-1513)
Ribo-Zero™ Magnetic Kit (Illumina, cat. no. MRZH116)
Stranded RNA-Seq Library Preparation Kit (KAPABiosystems, Roche, cat. no. 07962142001)
KAPA Library Quant Kit for Illumina Sequencing Platforms (Roche, cat. no. 07960140001).

Buffers:

Ribosomal purification:

Cell Lysis Buffer

20 mM HEPES-KOH (pH 7.4)

150 mM KCl

10 mM MgCl2

0.5mM DTT

rRNasin

RNaseOUT

SUPERaseIN

1 X cOmplete-EDTA-free protease inhibitor

High Salt Buffer

20 mM HEPES-KOH (pH 7.4)

350 mM KCl

10 mM MgCl2

0.5mM DTT

rRNasin

RNaseOUT

SUPERaseIN

100 µg/mL cycloheximide

1% IGEPAL CA-630

Nuclear purification:

Hypotonic Buffer H

20 mM HEPES, pH 7.9

1.5 mM $MgCl_2$

10 mM KCl

1 mM DTT

1X Complete protease inhibitor

supplemented with 0.01% Tricaine

Nuclear Purification Buffer (NPB)

10 mM HEPES, pH 7.9

40 mM NaCl

90 mM KCl
0.5 mM EDTA
0.5 mM spermidine
0.15 mM spermine
1 mM DTT
1X Complete protease inhibitor

NPBt
10 mM HEPES, pH 7.9
40 mM NaCl
90 mM KCl
0.5 mM EDTA
0.5 mM spermidine
0.15 mM spermine
1 mM DTT
1X Complete protease inhibitor
0.1% Triton X-100

Streptavidin bead preparation:

Solution A
0.1M NaOH, DEPC-treated
0.05M NaCl

Solution B
0.01M NaCl, DEPC-treated

Acknowledgements

This work was supported by March of Dimes Basil O'Connor Award, Lister Institute prize and BHF CRE Award to TSS and Clarendon Fund Fellowship to VCM.

Figure Legends:

Fig. 1. Biotagging toolkit for subcellular profiling in zebrafish.

Schematic of the binary transgenic system for cell-type specific *in vivo* biotinylation. BirA drivers in red, Avi effector lines in blue, POI, protein of interest. Cell-type and cell compartment specific isolation of RNA from double transgenic zebrafish embryos can be accomplished by Avi-tagging (blue) polyribosomes (top) or nuclei (bottom) after cell lysis and affinity binding with streptavidin magnetic beads. Polyribosome and nuclear transcriptome provides different transcript information, with bidirectional transcription detectable in the nuclear transcriptome.

Fig. 2: Bioanalyzer profiles as quality control measure for isolation of specific subcellular compartments.

Bioanalyzer profile of total RNA extracted from streptavidin-bound biotagged polyribosomes (A), nuclei (B), and whole embryo (C). The distinct profiles for polyribosomes and nuclei can be used to monitor the complete lysis of cells and isolation of the two subcellular compartments.

Reference:

1. Trask, H.W., et al., *Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs.* RNA, 2009. **15**(10): p. 1917-28.
2. Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* Science, 2008. **322**(5909): p. 1845-8.
3. Djebali, S., et al., *Landscape of transcription in human cells.* Nature, 2012. **489**(7414): p. 101-8.
4. Kaikkonen, M.U., M.T. Lam, and C.K. Glass, *Non-coding RNAs as regulators of gene expression and epigenetics.* Cardiovasc Res, 2011. **90**(3): p. 430-40.
5. Tollervey, D. and T. Kiss, *Function and synthesis of small nucleolar RNAs.* Curr Opin Cell Biol, 1997. **9**(3): p. 337-42.
6. Hamilton, A.J. and D.C. Baulcombe, *A species of small antisense RNA in posttranscriptional gene silencing in plants.* Science, 1999. **286**(5441): p. 950-2.
7. Iwasaki, Y.W., M.C. Siomi, and H. Siomi, *PIWI-Interacting RNA: Its Biogenesis and Functions.* Annu Rev Biochem, 2015. **84**: p. 405-33.
8. Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis.* Nat Rev Mol Cell Biol, 2014. **15**(8): p. 509-24.
9. Katayama, S., et al., *Antisense transcription in the mammalian transcriptome.* Science, 2005. **309**(5740): p. 1564-6.
10. Carninci, P., et al., *The transcriptional landscape of the mammalian genome.* Science, 2005. **309**(5740): p. 1559-63.
11. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome Res, 2012. **22**(9): p. 1775-89.
12. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.
13. Rinn, J.L. and H.Y. Chang, *Genome regulation by long noncoding RNAs.* Annu Rev Biochem, 2012. **81**: p. 145-66.
14. Kung, J.T., D. Colognori, and J.T. Lee, *Long noncoding RNAs: past, present, and future.* Genetics, 2013. **193**(3): p. 651-69.
15. Jeck, W.R., et al., *Circular RNAs are abundant, conserved, and associated with ALU repeats.* RNA, 2013. **19**(2): p. 141-57.
16. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency.* Nature, 2013. **495**(7441): p. 333-8.
17. Core, L.J., et al., *Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers.* Nat Genet, 2014. **46**(12): p. 1311-20.
18. Andersson, R., *Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model.* Bioessays, 2015. **37**(3): p. 314-23.
19. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-461.
20. Kim, T.K., et al., *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-7.
21. Preker, P., et al., *RNA exosome depletion reveals transcription upstream of active human promoters.* Science, 2008. **322**(5909): p. 1851-4.

22. Deal, R.B. and S. Henikoff, *A simple method for gene expression and chromatin profiling of individual cell types within a tissue.* Dev Cell, 2010. **18**(6): p. 1030-40.
23. Ooi, S.L., J.G. Henikoff, and S. Henikoff, *A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans.* Nucleic Acids Res, 2010. **38**(4): p. e26.
24. Steiner, F.A., et al., *Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling.* Genome Res, 2012. **22**(4): p. 766-77.
25. Amin, N.M., et al., *Proteomic profiling of cardiac tissue by isolation of nuclei tagged in specific cell types (INTACT).* Development, 2014. **141**(4): p. 962-73.
26. Housley, M.P., et al., *Translational profiling through biotinylation of tagged ribosomes in zebrafish.* Development, 2014. **141**(20): p. 3988-93.
27. de Boer, E., et al., *Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7480-5.
28. Hendrickson, W.A., et al., *Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation.* Proc Natl Acad Sci U S A, 1989. **86**(7): p. 2190-4.
29. Weber, P.C., et al., *Structural origins of high-affinity biotin binding to streptavidin.* Science, 1989. **243**(4887): p. 85-8.
30. Kim, J., et al., *Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells.* Nat Protoc, 2009. **4**(4): p. 506-17.
31. Trinh, L.A., et al., *Biotagging of Specific Cell Populations in Zebrafish Reveals Gene Regulatory Logic Encoded in the Nuclear Transcriptome.* Cell Rep, 2017. **19**(2): p. 425-440.
32. Heiman, M., et al., *Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP).* Nat Protoc, 2014. **9**(6): p. 1282-91.
33. Matunis, M.J., J. Wu, and G. Blobel, *SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex.* J Cell Biol, 1998. **140**(3): p. 499-509.
34. Hopper, A.K., H.M. Traglia, and R.W. Dunst, *The yeast RNA1 gene product necessary for RNA processing is located in the cytosol and apparently excluded from the nucleus.* J Cell Biol, 1990. **111**(2): p. 309-21.
35. Mahajan, R., et al., *A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2.* Cell, 1997. **88**(1): p. 97-107.
36. Rose, A. and I. Meier, *A domain unique to plant RanGAP is responsible for its targeting to the plant nuclear rim.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 15377-82.
37. Barthelson, R.A., et al., *Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells.* BMC Genomics, 2007. **8**: p. 340.
38. De Santa, F., et al., *A large fraction of extragenic RNA pol II transcription sites overlap enhancers.* PLoS Biol, 2010. **8**(5): p. e1000384.
39. Kowalczyk, M.S., et al., *Intragenic enhancers act as alternative promoters.* Mol Cell, 2012. **45**(4): p. 447-58.
40. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-61.
41. Deal, R.B. and S. Henikoff, *The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana.* Nat Protoc, 2011. **6**(1): p. 56-68.

42. Buranda, T., et al., *Peptides, antibodies, and FRET on beads in flow cytometry: A model system using fluoresceinated and biotinylated beta-endorphin.* Cytometry, 1999. **37**(1): p. 21-31.
43. Ingaramo, M. and D. Beckett, *Biotinylation, a post-translational modification controlled by the rate of protein-protein association.* J Biol Chem, 2011. **286**(15): p. 13071-8.
44. Kenyon, A., et al., *Active nuclear transcriptome analysis reveals inflammasome-dependent mechanism for early neutrophil response to Mycobacterium marinum.* Sci Rep, 2017. **7**(1): p. 6505.
45. Kenyon, A., et al., *Generation of a double binary transgenic zebrafish model to study myeloid gene regulation in response to oncogene activation in melanocytes.* Dis Model Mech, 2018. **11**(4).
46. Gavriouchkina, D., et al., *From pioneer to repressor: Bimodal foxd3 activity dynamically remodels neural crest regulatory landscape in vivo.* bioRxiv, 2017.
47. Goldman, J.A., et al., *Resolving Heart Regeneration by Replacement Histone Profiling.* Dev Cell, 2017. **40**(4): p. 392-404 e5.

Fig. 1

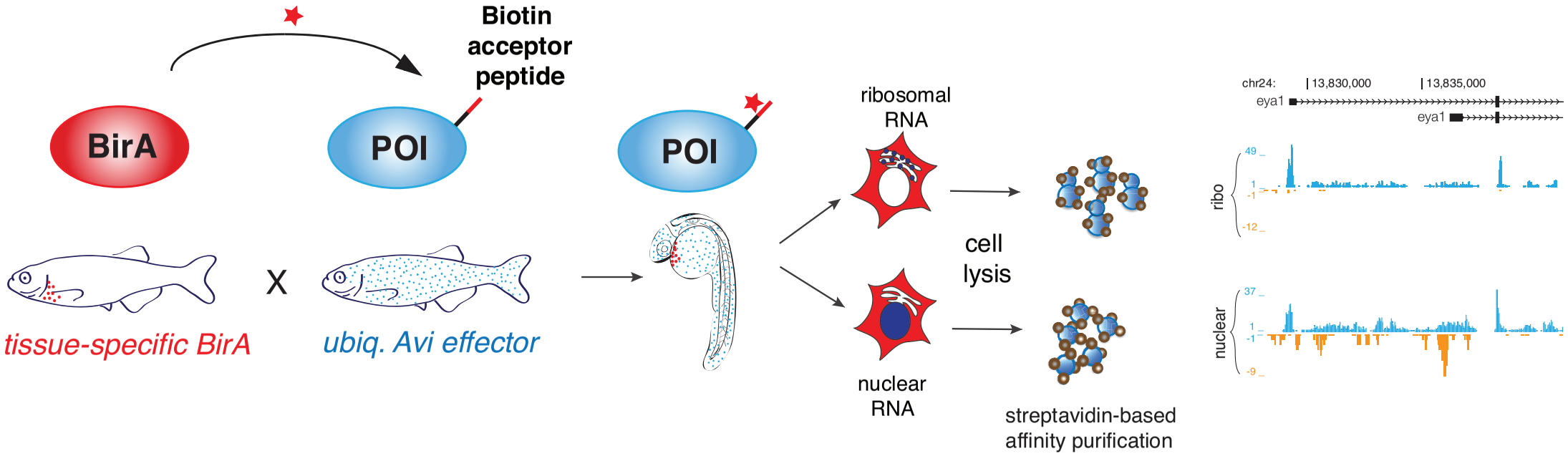
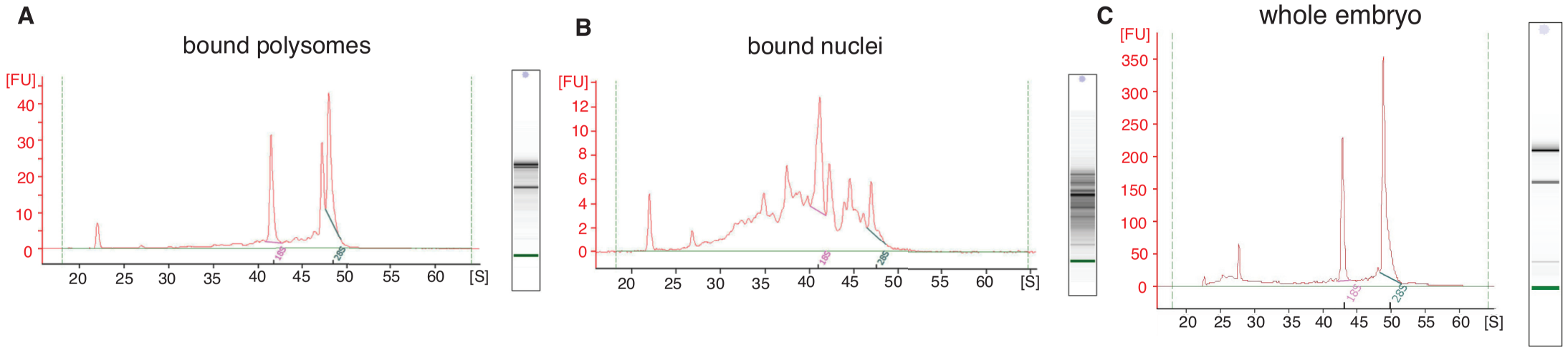

Biotin
acceptor
peptide
BirA
POI
POI
X
tissue-specific BirA
ubiq. Avi effector
ribosomal
RNA
nuclear
RNA
cell
lysis
streptavidin-based
affinity purification
chr24:
13,830,000
13,835,000
eya1
eya1
ribo
nuclear
49
1
-1
-12
37
1
-1
-9

Fig. 2


A
bound polysomes
[FU]
40
30
20
10
0
20 25 30 35 40 45 50 55 60 [S]
18S
28S
B
bound nuclei
[FU]
12
10
8
6
4
2
0
20 25 30 35 40 45 50 55 60 [S]
18S
28S
C
whole embryo
[FU]
350
300
250
200
150
100
50
0
20 25 30 35 40 45 50 55 60 [S]
18S
28S